S-E-C-R-E-T

T/III.I/R-1
19 November 1965

U N I T E D S T A T E S I N T E L L I G E N C E B O A R D

COMMITTEE ON DOCUMENTATION

TASK TEAM III - FOREIGN PUBLICATIONS

WORKING GROUP ON TRANSLITERATION

REPORT

1. A Working Group on Transliteration was established during the Fifth Meeting, 25 February 1965, of Task Team III, CODIB.

2. The Terms of Reference (Tab A) set the work objective of this group: To determine whether standardized systems of transliteration for foreign languages are needed within the Intelligence Community. Additionally, the group was directed to "consider the impact of transliteration in collection, processing, retrieval and production areas with special reference to its effect on systems of automation."

3. Members and Alternates:*

25X1

CIA	-	[redacted] Chairman
DIA	-	[redacted]
NSA	-	[redacted]

25X1

ARMY	-	Mr. Henry Holz
NAVY	-	Mr. P. Thomas Koines Mr. D. C. Ashworth
AIR FORCE	-	Mr. Wieslaw Arlet, ATD-LoC

25X1

CSS	-	[redacted] Secretary

*STATE was not represented and the member from DIA did not participate after the first meeting.

Other Contributors were:

25X1 NSA - [redacted]

AMS - Mr. Frank Shepard

25X1 CIA - [redacted]

AIR FORCE - Maj. Augustus Paine, FTD

4. Approach:

The working group assigned to investigate the need for standardization of transliteration systems within the Community was heterogenous with respect to background and working environment. Membership included linguists, analysts, EDP specialists in various combinations or lack of combination. It was recognized that "transliteration" though not quite a semantic stumbling block, implied a spectrum of characteristics each of which was valued by members in terms of familiar functional requirements. Because of this natural diversity of emphasis in individual viewpoint, it was necessary to uncover a channel of communication. To do this, we decided to hang our "search for need" on a study of the transliteration problems pertinent to a particular language, Russian. This language was included in the experience of a majority of members. Hopefully, those who had experience with a specific transliteration problem would stimulate those who lacked such experience into appreciation of other dimensions.

5. Open forum quickly exposed two points:

a. Standardization of transliteration could not be considered independently of the detailed parameters of such standardization.

b. Two main centers of gravity exist--machine and non-machine oriented, each positioned on a different but overlapping minimum set of characteristics as requirements.

6. The following was adopted as a non-inclusive working list of logical and linguistic characteristics for consideration in evaluating a transliteration system:

a. Non-ambiguity: A one to one correspondence between the elements of the source alphabet and the target alphabet. This implies reversibility between source and target elements.

b. Ambiguity:*

(1) Many to one - Two or more different elements in source alphabet have the identical element as representative of the target alphabet.

(2) One to many - One element in source language is represented by two or more distinct elements in the target language.

(3) One to n-graph - One element in source alphabet is represented by a combination of elements in target alphabet, each of which is a single target element which linguistically can occur with the other(s) juxtaposed in the same order. There are more complex variants wherein target alphabet "elements", in the context of this paragraph, are grouped single elements: Щ = SHCH

c. Phonetic readibility: Essentially desirable to a majority of readers. Pronunciation of target elements as target language phonemes will phonetically approximate pronunciation of original source language.

d. Can be accommodated by standard printing and input/output devices (computer printers, keypunches, typewriters, flexopunchers, page readers, etc.) with a minimum of modification (to the devices).

7. The minimum requirements of the machine oriented group is 6.a. with as much of 6.c. and 6.d. as is feasible. The significance of 6.a. in the machine environment is that it is efficient, economical and resolves a multiplicity of logical and technical problems. File conversion, language reversibility, valid file search and match, ease of data preparation, minimum computer programming, effective use of page-reading devices, etc., are all included as implications of non-ambiguous transliteration systems.

8. The non-machine oriented group subscribes to 6.c. as a minimum requirement and includes a degraded 6.a. by established practice. By this is meant that ambiguity is acceptable when its probability of occurring linguistically is very small. It must be emphasized, however, that a rationale based on "rarity of occurrence" would be academic at the first instance of a critical error in intelligence due to such ambiguity. All the major Russian systems BGN, LOC, BSI, and the old NSA system accept a few ambiguities. These systems are well established. A large volume of

*Non-reversibility is implied.

data has been published and filed in these systems. BGN has been extensively used within the Community. It is understandable that to date there has been reluctance by producers and users of these systems to change--since they have experienced little pain. But advances in machine processing and growth in volume of foreign language source data during the last decade is forcing a reappraisal directed either to the immediate necessities or to an estimate of future requirements.

9. There is evidence that the first pain threshhold has been exceeded at NSA, which might reasonably be anticipated given the existence there of a long-established and strong machine base. NSA has proposed the adoption of a new, Cyrillic-Roman transliteration system which in the opinion of this working group merits full CODIB attention as a suggested basis if and when it is decided to implement adoption of a standard system. Of great interest is a late arrival, Prof. R. O. Jakobson's (Harvard University) system as described in his article translated from the Russian, "On the Romanization of International Telegrams in Russian." The NSA and Jakobson systems are included in Tabs B and C as examples of transliteration systems which include the minimum requirements for automation.

10. Findings of the working group:

a. There is a need for non-ambiguous standard transliteration systems wherever machine files are exchanged, combined, searched for common entries, or transmitted. Minimally, the files can follow a different system as long as each system is one-to-one (machine conversion is then logically possible and technically simple).

b. A requirement exists for standardization of transliteration systems for the analyst-consumer who maintains a file or index with diverse inputs. Even though manual in character, his file is uneconomical because of the work required to convert all inputs into a consistent transliteration system. However, there is a less stringent requirement for absolute non-ambiguity since human judgment can operate.

c. No printed target alphabet element should be made up of more than two single alphabetic symbols. This is the maximum which will fit on a print pallet while maintaining readibility. It should again be emphasized that machine standardization will accommodate local differences with respect to choice of special printed characters as long as one-to-one representation is maintained.

d. A requirement for reversibility exists in the Community, but in special areas of work.

e. Mr. Shephard of AMS pointed out that BGN Cyrillic-Roman was established in 1947 in response to requirements for transliteration of place names for mapping purposes. It was designed to provide a pronounceable Roman form for Cyrillic place names. Automation problems were not a consideration at that time. Mr. Shephard suggested that the BGN would be interested in studying any proposed standard for the Community.

f. Production (consumption) areas in the CIA generally approve of the desirability of standardization. However, they consider that the use of BGN, which is well established, allows them well manageable problems, and looks like a proper candidate for a standard. These areas have had experience mainly with manual and EAM files. The exception to this view is found in the Office of Computer Services which is conscious of the projected growth in automatic information processing and its implications.

g. Mr. Holz, Army, reported that only those working in the machine environment in his Department appeared to be aware of a problem in handling ambiguous transliteration systems. He noted trends indicating greater use of computers which would force the use of non-ambiguous systems.

h. Navy's views were expressed by Mr. Koines as follows:

(1) Transliteration presents no difficulty at all to the users of foreign language (primarily Russian, some Chinese) materials. The analysts are able to adapt to the system of transliteration being used; when they have a problem, they turn to the ONI Translation Section for assistance.

(2) The Navy Scientific and Technical Intelligence Center (STIC) is in the advanced planning stage of converting its scientific and technical information files to punched cards. There is no target date for completion of the conversion. STIC then expects to put its information on tapes in the likewise indefinite, and more remote, future. Cognizant STIC personnel agree, however, that a transliteration system for use in a computer must be unambiguous and standardized, with a view to possible exchanges of data between agencies. They express no special preference of data for any transliteration system for daily use, as long as it is pronounceable. This excludes the use of any letters other than Roman.

(3) Part of the Maritime Intelligence Branch has data on merchant ships and ship movements on tape. The spelling (transliteration) of the ship name is only one of the three parameters used in identifying ships; consequently, transliteration poses no appreciable problem to that Branch.

i. Mr. Arlet, Air Force, stated that the BGN system* satisfies the Air Force requirement. The full text of this position, which includes a point by point critique of the NSA suggested system, and the NSA reply thereto, are attached as Tab D without comment.

11. Conclusions and Recommendations:

a. A non-ambiguous Cyrillic-Roman transliteration system, which includes both machine code and graphic standards, should be developed for early use by those groups within the Community which process or maintain common files.

b. The graphic standard referred to in 11.a. cannot be any of the established systems which contain ambiguities as defined in 6.b. It is here that compromises must be made with respect to established practice, linguistic esthetics, and the demands of printing and reading devices.

c. Resolution of the Cyrillic-Roman transliteration problem within the Community should establish both a precedent and guide for using the same approach to the transliteration of other languages which are problems with respect to volume and machine requirements.

d. The adoption of a transliteration standard implies partial incompatibility with older files maintained in established systems. Although the problem has not been within the purview of this group, we do suggest an approach. It is clear that total conversion of old files to a new standard is not a manageable task. Therefore, new files should be built while maintaining and converting parts of the old files as conditions allow. In any case, the impossibility of total assimilation, and inconvenience in the temporary existence of two sets of files should not weight the CODIB decision against implementation of a standard. While the present need is evident, it is imperative that we prepare for the future day when there will exist both the

*BGN system of identifying and officially naming places.

requirement and capability to process and store masses of data in a larger order of magnitude than at the present time.

25X1

Chairman, Transliteration Working Group